19th International Conference on Environmental Science and Technology

Kos, Greece, 3 to 6 September 2025

Bio-oil production from biomass pyrolysis via a hybrid mathematical and machine learning approach

XU F. [1], LYONS CERON A. [1], NESHUMAJEV D. [1], KONIST A. [1,*]

[1] Department of Energy Technology, Tallinn University of Technology, Tallinn, 19086, Estonia

*corresponding author:
e-mail: alar.konist@taltech.ee

Abstract: A hybrid mathematical and machine learning (HMML) approach is applied for maximizing bio-oil yield during the continuous biomass pyrolysis, using local common reed as feedstock. The HMML approach, which combines Response Surface Methodology (RSM) and Machine Learning algorithms (Gaussian Process (GPR) and Support Vector Regression (SVR)), provides a powerful tool for experimental optimization. Temperature, carrier gas flow rate, and feeding rate were the major parameters studied and optimized. The HMML approach demonstrated a strong predictive capability, as indicated by the statistical results. The optimal conditions (490 °C, 220 SCCM, 10.8 RPM) were identified and implemented, achieving a validated bio-oil yield of 39.90 ± 0.19 wt.%. The results confirm HMML as an effective tool for optimizing complex pyrolysis processes and improving bio-oil yield.

Keywords: Pyrolysis, Response surface methodology, Machine learning, Bio-oil, Process optimization.

1. Introduction

Bio-oil is one of the promising alternatives to fossil fuels through the liquification (e.g., pyrolysis) process of biomass and waste. Operation parameters, e.g., temperature, feeding rate, etc., play an essential role in the quality and quantity of products during pyrolysis. The yield and composition of bio-oil are associated with the operating parameters, in addition to the source of biomass (Carpenter *et al.*, 2014).

Hence, process optimization is important for determining the optimal conditions for bio-oil production in biomass pyrolysis. Previous studies have summarized that temperature has a significant impact on bio-oil yield, with certain effects caused by other parameters, such as heating rate, gas flow rate, residence time, etc. (Guedes *et al.*, 2018). Traditional methods in process optimization, e.g., one factor at a time, are simple and intuitive, but limited in detecting interactions and optimizing complex systems. Some mathematical methods, e.g., RSM, can be applied for further optimization and multivariable interaction analysis (Veza *et al.*, 2023). In addition, machine learning (ML) approaches are also introduced into biomass conversion studies to predict product yields, decomposition behaviors, and kinetic parameters (Ascher *et al.*, 2022). Since the RSM can only capture and implement second-order relationships, the combination with ML could improve the accuracy of predictions.

2. Materials and Methods

2.1. Biomass pyrolysis experiments

Local common reed (*Phragmites australis*) was selected as the biomass feedstock for bio-oil production through fast pyrolysis (Xu *et al.*, 2025). The continuous pyrolysis facility, consisting of a cylindrical retort equipped with feeding, heating, and condensing units, was employed for the biomass pyrolysis experiments (Ceron *et al.*, 2024). Three parameters, including temperature (400-550 °C), carrier gas flow rate (150-350 SCCM), and feeding rate (5-15 RPM), were considered during pyrolysis for process optimization.

2.2. Hybrid mathematical and machine learning approach

A Central Composite Design (CCD) was applied for the design of experiments, which resulted in 18 initial runs based on 3 factors and 5 levels. The results were processed using RSM in conjunction with ML, also called a hybrid mathematical and machine learning (HMML) approach. It is a predictable and interpretable model that maximizes bio-oil yield via parameter optimization. The optimal conditions were predicted using the HMML approach and were validated to obtain the actual bio-oil yield. Products produced under optimal conditions were characterized according to related standards and infrastructures. The detailed workflow for the HMML enhanced bio-oil production process is presented in **Figure 1(a)**.

3. Results and Discussion

3.*1. Process optimization via HMML*

The contour plots, as presented in **Figure 1(b, c, d)**, illustrate the influence of variable interactions on bio-oil yield. The contour plot of temperature versus flow rate (**Figure 1b**) reveals an elliptical response surface where bio-oil yield is maximized. Similarly, the interaction between temperature and feeding rate (**Figure 1c**) shows a curved contour pattern, which is also similar to the relationship between flow rate and feeding rate (**Figure 1d**). These plots visually confirm the existence of an optimal operating condition across all three parameters. Furthermore, the full quadratic model exhibited a high degree of fit (**Figure 1e**), with an adjusted R^2 of 0.9789, which indicates an excellent agreement between experimental and predicted values.

Two algorithms, Gaussian Process (GPR) and Support Vector Regression (SVR), were trained and evaluated via hyperparameter tuning to achieve highly accurate predictions and assess model generalization. **Figure 1f** and **1g** show that the GPR and SVR models exhibited great performance, with high R^2 and low mean absolute error (MAE) values. For cross-validation, the two models also maintained a good generalization capability from the evaluation. Overall, both models were capable of capturing the nonlinear relationships between process variables and bio-oil yield.

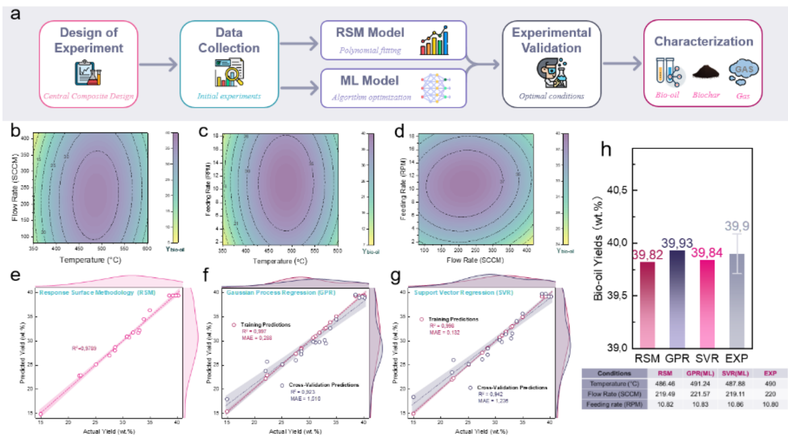
3.2. Bio-oil production

Through the HMML approach, the optimal condition and maximum yield during reed pyrolysis were predicted as shown in **Figure 1h**. For example, the maximum yield of bio-oil can be produced at a temperature of 486.46 °C, a flow rate of 219.49 SCCM, and a feeding rate of 10.82 RPM, according to the predicted results via RSM. Similar parameters were predicted by ML algorithms (GPR and SVR) as well, as shown in **Figure 1h**. The close match indicates that the HMML approach by integrating RSM and ML agrees on the trends and optimum points, which boosts confidence in the results. Considering the actual implementation, the specific optimal experimental condition set is 490°C, 220 SCCM, and 10.8 RPM.

The predicted bio-oil yields are also closely aligned across the models, i.e., 39.82 wt.% for RSM, 39.93 wt.% for GPR, and 39.84 wt.% for SVR, respectively. It also matches the actual bio-oil yield conducted under optimal conditions, resulting in 39.90 ± 0.19 wt.%. The yielded bio-oil via HMML optimization is slightly higher than in previous studies using reed straw (36.52 wt.%) *(Hu* et al.*, 2023)*. The characteristics of bio-oil will be analyzed and discussed later.

4. Conclusions

This study successfully applied a hybrid RSM and ML approach to optimize fast pyrolysis conditions for bio-oil production from reed biomass. The HMML model accurately predicted optimal parameters, which yield bio-oil at 39.90 ± 0.19 wt.% that enhances bio-oil production from reed. The method offers a reliable, interpretable framework for optimizing bio-oil production and advancing biomass conversion technologies.



Conditions	RSM	GPR(ML)	SVR(ML)	EXP
Temperature (°C)	486.46	491.24	487.88	490
Flow Rate (SCCM)	219.49	221.57	219.11	220
Feeding rate (RPM)	10.82	10.83	10.86	10.80

Figure 1. The hybrid mathematical and machine learning (HMML) approach workflow (a), contour plots between variables from RSM analysis (b-d), comparison between predicted and actual bio-oil yield via HMML (e-g), and predicted bio-oil yields from HMML and actual bio-oil under optimal conditions (h).

Acknowledgement

The authors greatly appreciated the funding from the Estonian Research Council (TEM-TA73) and the European Union Cohesion and Internal Security Policy Fund.

References

Ascher, S., Watson, I., You, S. (2022). Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. *Renewable and Sustainable Energy Reviews*, **155**, 111902

Carpenter, D., Westover, T.L., Czernik, S., et al. (2014). Biomass feedstocks for renewable fuel production: A review of the impacts of feedstock and pretreatment on the yield and product distribution of fast pyrolysis bio-oils and vapors. *Green Chemistry*, **16**, 384–406

Ceron, A.L., Pihu, T., Konist, A. (2024). Co-pyrolysis of biomass woodchips with Ca-rich oil shale fuel in a continuous feed reactor. *Oil Shale*, **41**, 208

Guedes, R.E., Luna, A.S., Torres, A.R. (2018). Operating parameters for bio-oil production in biomass pyrolysis: A review. *Journal of Analytical and Applied Pyrolysis*, **129**, 134–49

Hu, J., Chen, L., Chen, P., et al. (2023). One-pot cogeneration of phenol-rich bio-oil, hydrogen-rich gas and solid carbon degradation material from reed. *Fuel Processing Technology*, **250**, 107912

Veza, I., Spraggon, M., Fattah, I.M.R., et al. (2023). Response surface methodology (RSM) for optimizing engine performance and emissions fueled with biofuel: Review of RSM for sustainability energy transition. *Results in Engineering*, **18**, 101213

Xu, F., Nešumajev, D., Konist, A. (2025). Thermal decomposition behaviors and kinetic parameter calculations during common reed and its components pyrolysis. *Renewable Energy*, **248**, 123130